Intonation and Computation: Deception

Julia Hirschberg
LSA 2017
julia@cs.columbia.edu
Deceptive Speech: A Case Study

• How do you come up with an idea for an empirical speech study?
• How do you go about doing it?
• What problems might you run into along the way?
Deception & Culture

- Not widely researched

- What has been looked at:
  - How culture affects when people deceive (e.g., Lapinski & Levine, 2000; Seiter et al, 2002)
  - Beliefs about deceptive behaviors across cultures
    - But contrasting findings (e.g., Al-Simadi, 2000; Bond & The Global Deception Research Team 2006)

- Few studies on universal versus culture-specific verbal cues (e.g., Cody et al. 1989; Cheng & Broadhurst, 2005)
  - But have examined one culture per study
Defining Deception

• Deliberate choice to mislead
  – Without prior notification
  – To gain some advantage or to avoid some penalty

• Not:
  – Self-deception, delusion, pathological behavior
  – Theater
  – Falsehoods due to ignorance/error

• Everyday (white) Lies hard to detect

• But Serious Lies?
Why are ‘serious’ lies difficult?

• Hypotheses:
  – Our cognitive load is increased when lying because…
    • Must keep story straight
    • Must remember what we’ve said and what we haven’t said
  – Our fear of detection is increased if…
    • We believe our target is hard to fool or suspicious
    • Stakes are high: serious rewards and/or punishments
• Makes it hard for us to control indicators of deception
Cues to Deception in Many Modalities

• **Body posture and gestures** (Burgoon et al‘94)
  – Complete shifts in posture, touching one’s face,…

• **Microexpressions** (Ekman‘76, Frank‘03)
  – Fleeting traces of fear, elation,…

• **Biometric factors** (Horvath‘73)
  – Increased blood pressure, perspiration, respiration, odor…other correlates of stress → polygraph

• **Changes in brain activity: true vs. false stories**

• **Variation in what is said and how** (Adam‘96, Pennebaker et al‘01, Newman et al ‘03, Streeter et al’77, Hirschberg et al ‘05, Levitan et al ’16)
Lying and LIWC (Newman et al ’03)

- **Linguistic Inquiry and Word Count:**
  - Do false (text) stories differ from true ones?
  - LIWC classifies 61-67% of stories in multiple topics
    - Context/topic did matter
  - Liars are less cognitively complex, use fewer self- or other-references, and use more negative emotion words
    - *Self-reference indicates ownership of statement*
    - *Negative emotion words indicate guilt*
    - *Creating false stories increases cognitive load*
  - LIWC can be useful but…
Current Approaches to Deception Detection

• Training humans
  – John Reid & Associates
    • Behavioral Analysis: Interview and Interrogation
      • *I didn’t take the money vs. I did not take the money*

• Laboratory studies: Production and Perception

• ‘Automatic’ methods
  – Polygraph
  – Nemesysco and the Love Detector
  – *No evidence that any of these work….but publishing this statement can be dangerous!* (Anders Eriksson and Francisco La Cerda)
Our Research

• Conduct *objective, experimentally verified* studies of cues to deception which predict better than *humans* or *polygraphs*

• *Our method:*
  – Collect speech data and extract *acoustic, prosodic, and lexical cues* automatically
  – Take *gender, ethnicity, and personality factors* into account
  – Use *Machine Learning* techniques to train models to classify deceptive vs. non-deceptive speech
Previous Work

- Created Columbia/SRI/Colorado Deception Corpus
  - Within-subject recordings of deceptive and non-deceptive speech
    - 32 adult native American English speakers
    - 25-50m interviews by trained interviewer (Reid technique): 15.2h of speech, 7h from subjects
    - Subjects given tasks and incentivized to lie about performance
    - Ground truth identified by subjects
Acoustic/Prosodic Features

- Duration features
  - Phone / Vowel / Syllable Durations
  - Normalized by Phone/Vowel Means, Speaker
- Speaking rate features (vowels/time)
- Pause features (cf Benus et al. ‘06)
  - Speech to pause ratio, number of long pauses
  - Maximum pause length
- Energy features (RMS energy)
- Pitch features
  - Pitch stylization (Sonmez et al. ‘98)
  - Model of F0 to estimate speaker range
  - Pitch ranges, slopes, locations of interest
- Spectral tilt features
Lexical Features

- Presence and # of filled pauses
- Is this a question? A question following a question
- Presence of pronouns (by person, case and number)
- A specific denial?
- Presence and # of cue phrases
- Presence of self repairs
- Presence of contractions
- Presence of positive/negative emotion words
- Verb tense
- Presence of ‘yes’, ‘no’, ‘not’, negative contractions
- Presence of ‘absolutely’, ‘really’

- Presence of hedges
- Complexity: syls/words
- Number of repeated words
- Punctuation type
- Length of unit (in sec and words)
- # words/unit length
- # of laughs
- # of audible breaths
- # of other speaker noise
- # of mispronounced words
- # of unintelligible words
Subject-Dependent Features

- % units with cue phrases
- % units with filled pauses
- % units with laughter
- Lies/truths with filled pauses ratio
- Lies/truths with cue phrases ratio
- Lies/truths with laughter ratio
- Gender
Results

• 88 features, normalized within-speaker
  – Discrete: Lexical, discourse, pause
  – Continuous features: Acoustic, prosodic, paralinguistic, lexical

• Best Performance: Best 39 features + c4.5 ML
  – Accuracy: 70.00%
  – TRUTH F-measure: 75.78
  – Lexical, subject-dependent & speaker-normalized features best predictors
  – Interesting individual differences: how to predict?
Evaluation: Compared to Human Deception Detection

• Most people are very poor at detecting deception
  – ~50% accuracy (Ekman & O’Sullivan ‘91, Aamodt ‘06)
  – People use unreliable cues, *even with training*

• This study
  – 32 Judges, rating 2 interviews
  – Received ‘training’ on one subject

• Pre- and post-test questionnaires

• Personality Inventory
A Meta-Study of Human Deception Detection  
(*Aamodt & Mitchell 2004*)

<table>
<thead>
<tr>
<th>Group</th>
<th>#Studies</th>
<th>#Subjects</th>
<th>Accuracy %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Criminals</td>
<td>1</td>
<td>52</td>
<td>65.40</td>
</tr>
<tr>
<td>Secret service</td>
<td>1</td>
<td>34</td>
<td>64.12</td>
</tr>
<tr>
<td>Psychologists</td>
<td>4</td>
<td>508</td>
<td>61.56</td>
</tr>
<tr>
<td>Judges</td>
<td>2</td>
<td>194</td>
<td>59.01</td>
</tr>
<tr>
<td>Cops</td>
<td>8</td>
<td>511</td>
<td>55.16</td>
</tr>
<tr>
<td>Federal officers</td>
<td>4</td>
<td>341</td>
<td>54.54</td>
</tr>
<tr>
<td>Students</td>
<td>122</td>
<td>8,876</td>
<td>54.20</td>
</tr>
<tr>
<td>Detectives</td>
<td>5</td>
<td>341</td>
<td>51.16</td>
</tr>
<tr>
<td>Parole officers</td>
<td>1</td>
<td>32</td>
<td>40.42</td>
</tr>
</tbody>
</table>
Table 1: Judges’ aggregate performance classifying TRUTH/LIE.

<table>
<thead>
<tr>
<th>Lie Category</th>
<th>Chance Baseline</th>
<th>Mean&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>63.87&lt;sup&gt;b&lt;/sup&gt;</td>
<td>58.23</td>
<td>57.42</td>
<td>7.51</td>
<td>40.64</td>
<td>71.48</td>
</tr>
<tr>
<td>Global</td>
<td>63.64&lt;sup&gt;c&lt;/sup&gt;</td>
<td>47.76</td>
<td>50.00</td>
<td>14.82</td>
<td>16.67</td>
<td>75.00</td>
</tr>
</tbody>
</table>

<sup>a</sup>Each judge’s score is his or her average over two interviews; as percentages.

<sup>b</sup>Guessing TRUTH each time.

<sup>c</sup>Guessing LIE each time.

---

Table 1: Aggregate performance by interviewee.

<table>
<thead>
<tr>
<th>Lie Type</th>
<th>Mean&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Median</th>
<th>Std. Dev.</th>
<th>Min.</th>
<th>Max.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Local</td>
<td>58.23</td>
<td>58.58</td>
<td>9.44</td>
<td>35.86</td>
<td>87.79</td>
</tr>
<tr>
<td>Global</td>
<td>44.83</td>
<td>45.58</td>
<td>17.40</td>
<td>10.00</td>
<td>81.67</td>
</tr>
</tbody>
</table>

<sup>a</sup>Each interviewee’s score is the average over two judges; as percentages.
What Makes Some People Better Judges?

• Does personality play a role?
The Big Five (Costa & McCrae, 1992)

- **Openness to Experience:** intellect/imagination, originality, curiosity, ingenuity, imaginative, insightful “I have a lot of intellectual curiosity.”

- **Conscientiousness:** orderliness, responsibility, dependability, organized, thorough, planful “I strive for excellence in everything I do.”

- **Extraversion:** talkativeness, assertiveness, energy “I like to have a lot of people around me.”
• **Neuroticism:** anxiety, emotional instability, intense, moody, “I often feel inferior to others.”

• **Agreeableness:** good-naturedness, cooperativeness, sympathetic, kind, affectionate, trust “I would rather cooperate with others than compete with them.”
Neuroticism, Openness & Agreeableness Correlate with Judge’s Performance

On Judging Global lies.

Table 1: Correlations between personality factors and judge performance at labeling global lies.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Measure</th>
<th>Pearson’s corr. coeff.</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neuroticism</td>
<td>Proportion of segments judged LIE</td>
<td>-0.44</td>
<td>0.012</td>
</tr>
<tr>
<td>Openness</td>
<td>Accuracy</td>
<td>0.51</td>
<td>0.003</td>
</tr>
<tr>
<td>Agreeableness</td>
<td></td>
<td>0.41</td>
<td>0.021</td>
</tr>
<tr>
<td>Neuroticism</td>
<td>F-measure for TRUTH</td>
<td>0.37</td>
<td>0.035</td>
</tr>
<tr>
<td>Agreeableness</td>
<td></td>
<td>0.41</td>
<td>0.019</td>
</tr>
<tr>
<td>Openness</td>
<td>F-measure for LIE</td>
<td>0.52</td>
<td>0.003</td>
</tr>
</tbody>
</table>
Current Study: Hypotheses

- **Personality factors** can help to predict differences in deceptive behavior.
- **Subjects who deceive better can also detect deception better.**
- **Cultural differences and gender also play a role in deceptive behavior and in deception detection abilities** (Cheng & Broadhurst 2005).
- **New task:** Studies of pairs of American English and Mandarin Chinese native speakers, speaking English, interviewing each other.
Cross-Cultural Deception Detection (Cheng & Broadhurst ‘05)

• Immigration motivation
  – Is it easier to judge deception in your own L1?
  – Do verbal and non-verbal cues differ for L1 and L2?

• 26 criminology students judged 20 video interviews in 2x2 betw subjects opinion paradigm (Cantonese/English bilinguals)
  – Better at judging lies but worse at judging truth in 2nd language
  – 2nd language affected subjects’ verbal and nonverbal behavior whether telling truth or lie
Our CxC Experiment
Our CxC Experiment
Our CxC Experiment
## Biographical Questionnaire

**Participant No. ______**  

**Date ________**

### Instructions

Please carefully look through the questions. Write down the true answer to each question in the "True Answer" column. When you have finished that, for all the questions that have don't have "X's in the "False Answer" column, make up an answer. Consult the additional sheet you have been given. You want to choose a lie that you are not as familiar with as the true answer.

This experiment is completely anonymous—your name will never be linked to the data.

<table>
<thead>
<tr>
<th>No.</th>
<th>Questions</th>
<th>True Answer</th>
<th>False Answer</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Where were you born?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>How many years did you live in your first home?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>What is your mother's job?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>What is your father's job?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Have your parents divorced?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>Have you ever broken a bone?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>Do you have allergies to any foods?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>Have you ever stayed overnight in a hospital as a patient?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>Have you ever tweeted? (posted a message on twitter)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>Have you ever bought anything on eBay?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>Do you own an e-reader of any kind?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>Who was the last person you were in a physical fight with?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>Have you ever gotten into trouble with the police?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>Who ended your last romantic relationship?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>Whom do you love more, your mother or father?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>What is the most you have ever spent on a pair of shoes?</td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>What is the last movie you saw that you really hated?</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Our CxC Experiment
Our CxC Experiment
Our CxC Experiment

Background survey
Biographical Questionnaire
Baseline sample
Lying game
Survey
Our CxC Experiment
Our CxC Experiment
Scoring and Motivation

• Success
  – Ability to lie -> as interviewee, number of lies believed true by interviewer
  – Ability to detect lies -> as interviewer, number of correct guesses for truth and lie

• Note: $1 added or subtracted for each right or wrong decision
Example: “Where were you born?”

True or False?
Example: “Where were you born?”

False!
Annotation

• Transcribed using Amazon Mechanical Turk
  – 5 Turkers per utterance, combined using Rover techniques

• Automatically segmented using Praat into Inter-Pausal Units (IPUs) at 50ms silence

• Automatically aligned with speech and truth/lie labels using aligner built with Kaldi
## Rover Example

<p>| | | | | | | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>its</td>
<td>really</td>
<td>fun</td>
<td>um</td>
<td>I</td>
<td>go</td>
<td>like</td>
<td>to</td>
<td>a place</td>
</tr>
<tr>
<td>2</td>
<td>its</td>
<td>really</td>
<td>fun</td>
<td>i</td>
<td>go</td>
<td>like</td>
<td>a</td>
<td>place</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>it’s</td>
<td>really</td>
<td>fun</td>
<td>um</td>
<td>I</td>
<td>go</td>
<td>like</td>
<td>to</td>
<td>a place</td>
</tr>
<tr>
<td>Rover output</td>
<td>its</td>
<td>really</td>
<td>fun</td>
<td>um</td>
<td>I</td>
<td>go</td>
<td>like</td>
<td>to</td>
<td>a place</td>
</tr>
<tr>
<td>score</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2/3</td>
<td>2/3</td>
<td>1</td>
<td>1</td>
<td>2/3</td>
<td>1</td>
</tr>
</tbody>
</table>

ROVER Score = \( \frac{1+1+1+2/3+2/3+1+1+2/3+1+1}{10} = 0.9 \)
Corpus

- Two current versions:
  - Balanced corpus: 140 pairs; 112h speech
  - Larger corpus: 172 pairs; 122h speech
  - Largest existing cleanly recorded corpus of deceptive/non-deceptive speech
Correlations: Deceptive Ability & Deception Detection

- People’s ability to detect deception correlates with their ability to deceive \( r(280) = 0.12, p = 0.05 \)
  - Ability to lie -> as interviewee, # lies believed true by interviewer
  - Ability to detect lies vs. truth -> as interviewer, # correct guesses for truth or lie

- Holds across all subjects but strongest for **females** \( r(140) = 0.24, p = 0.005 \)
Correlations: Gender, Ethnicity, Personality & Ability to Deceive

• *Extraversion* is significantly *negatively correlated* with ability to deceive for English males $r(70) = -0.24$, $p = 0.04$

• Tendencies for
  – Chinese/female *extraversion positively* correlated with ability to deceive
  – American/female *conscientiousness negatively* correlated with ability to deceive
Gender, Ethnicity, Personality & Deception Detection

• No effect of personality factors in deception detection found so far
  – Contra earlier findings for English speakers (Enos et al ’06)
  – But this is real-time detection vs. later judgments
  – Working on better ways representations
    • Clustering using all factors
    • Quantizing raw scores: high, medium, low
Confidence in Judgments

• Ability to detect deception negatively correlates with confidence in judgments for all subjects $r(278) = -0.15$, $p = 0.01$
  – Strongest for females $r(140) = -0.25$, $p = 0.003$

• Ability to deceive negatively correlated with confidence for males $r(138) = -0.17$, $p = .05$
  – Strongest for Chinese males $r(70) = -0.27$, $p = 0.02$

• Less confident interviewers may ask more follow-up questions and obtain more evidence for decisions?
Personality Factors and Confidence

• Neuroticism *negatively* correlates with confidence for Chinese female subjects $r(70) = -0.29$, $p = 0.02$

• Openness to experience *negatively* correlates with confidence for all subjects $r(277) = -0.14$, $p = 0.02$
  – Strongest for females $r(140) = -0.20$, $p = 0.02$
  – Strongest for Chinese females $r(70) = -0.29$, $p = 0.02$
Classification Results (Unbalanced Corpus)

• **Features** (z-score normalized):
  – Acoustic features: $f_0$, intensity, voice quality, speaking rate
    – raw and normalized 2 ways
  – Gender: subject and partner
  – Ethnicity: subject and partner
  – Personality scores
  – Lexical features not yet available for entire corpus

• **Weka experiments**
  – J48 decision trees
  – Random Forests
  – Bagging
Classification Results

<table>
<thead>
<tr>
<th>Model</th>
<th>Raw</th>
<th>SessionNorm</th>
<th>BaselineNorm</th>
</tr>
</thead>
<tbody>
<tr>
<td>J48</td>
<td>59.89</td>
<td>62.09</td>
<td>62.19</td>
</tr>
<tr>
<td>Bagging</td>
<td>58.65</td>
<td>61.19</td>
<td>61.01</td>
</tr>
<tr>
<td>RF</td>
<td>61.23</td>
<td><strong>63.03</strong></td>
<td>62.79</td>
</tr>
</tbody>
</table>

- Baseline accuracy: 59.9%
Added Features

- Speaker gender
- Speaker native language
- NEO-FFI personality scores
## Adding Gender, Ethnicity and Personality Features (SessionNorm)

<table>
<thead>
<tr>
<th>Model</th>
<th>Acoustic/prosodic</th>
<th>Acoustic/prosodic +gender/ethn/NEO</th>
</tr>
</thead>
<tbody>
<tr>
<td>J48</td>
<td>62.09</td>
<td>64.86</td>
</tr>
<tr>
<td>Bagging</td>
<td>61.19</td>
<td>63.9</td>
</tr>
<tr>
<td>RF</td>
<td>63.03</td>
<td>65.86</td>
</tr>
</tbody>
</table>

- Majority class baseline accuracy: 59.9%
Other Questions

• Once you have a corpus….what other questions can you investigate?
Interviewer Correctness and Confidence by Question

![Confidence Score vs Question Number](image-url)
Which Question Types are Most Difficult?

• Yes/no vs. open-ended questions
  – Do you own an e-reader of any kind?
  – Who ended your last romantic relationship?

• Sensitive (intrusive, threat of disclosure, social desirability) vs. non-sensitive
  – Who ended your last romantic relationship?
  – Do you own an e-reader of any kind?
Current Findings

• All interviewers better at guessing T/F for yes/no questions than for open-ended questions
  – Only English interviewers detected lies to ynq’s better than to open-ended questions
  – Many differences in pair types wrt detecting lies to ynq’s (e.g. English females did better when interviewing Mandarin males than English males)

• All interviewers detected T/F better in sensitive questions than non-sensitive, with some differences for pair types
Baseline Interviews & Personality Predictions

• 3-4m pre-experiment interviews between subject and experimenter to use for normalizing interview data

• Can we detect gender, ethnicity, personality types and even ability to lie or detect lies from this data alone?
Baseline Interview

- Tell me how you decided to come to Columbia.
- What are your favorite things to do in New York City?
- What do you like the most about living in New York City?
- What do you like the least about living in New York City?
- Describe a typical weekend for you, from Friday night through Sunday night.
- What was the best food you ever ate, where did you have it, and what made it so good?
- Where was the last place you traveled, and what are some things you did while you were there?
- What was the last movie you saw, and what was the premise of it?
- Besides work or school, what do you do with your time?
- What did you do this past summer?
Findings

• 3-4 minutes of ’normal’ speech turn out to yield good classification of ethnicity, gender, and success at deception detection, with some success at predicting NEO-FFI personality scores as well

• Promising results for future work on classifying speakers from small amounts of ‘normal’ speech
Gender Classification

- Baseline: 61% female
- f0 is the best feature
- Useful LIWC categories: ‘anger’, ‘money’, ‘certainty’ -> male

<table>
<thead>
<tr>
<th>Features</th>
<th>acc</th>
<th>prec</th>
<th>rec</th>
<th>f-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosodic</td>
<td>.94</td>
<td>.94</td>
<td>.96</td>
<td>.97</td>
</tr>
<tr>
<td>LIWC</td>
<td>.70</td>
<td>.69</td>
<td>.86</td>
<td>.76</td>
</tr>
<tr>
<td>Combined</td>
<td>.94</td>
<td>.94</td>
<td>.98</td>
<td>.96</td>
</tr>
</tbody>
</table>

Native Language

- Baseline: 57% English
- Useful LIWC categories: ‘apostrophe’ -> English (contractions)
  ‘dash’ -> Mandarin (false starts)

<table>
<thead>
<tr>
<th>Features</th>
<th>acc</th>
<th>prec</th>
<th>rec</th>
<th>f-score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosodic</td>
<td>.65</td>
<td>.64</td>
<td>.40</td>
<td>.52</td>
</tr>
<tr>
<td>LIWC</td>
<td>.84</td>
<td>.82</td>
<td>.7</td>
<td>.76</td>
</tr>
<tr>
<td>Combined</td>
<td>.81</td>
<td>.88</td>
<td>.68</td>
<td>.78</td>
</tr>
</tbody>
</table>
Personality Classification (F measure) (f1)measures)

- Binning: high, medium, low for each factor, based on population means
- Compare AdaBoost and RandomForest classifiers
- Useful LIWC categories:
  - Neuroticism: ‘power’, ‘money’
  - Extraversion: ‘drives’, ‘focusfuture’
  - Openness to experience: ‘interrogation’, ‘focuspast’
  - Agreeableness: ‘social’, ‘assent’
  - Conscientiousness: ‘work’, ‘time’

<table>
<thead>
<tr>
<th>Features</th>
<th>Model</th>
<th>N</th>
<th>E</th>
<th>O</th>
<th>A</th>
<th>C</th>
</tr>
</thead>
<tbody>
<tr>
<td>Prosodic</td>
<td>AB</td>
<td>.43</td>
<td>.30</td>
<td>.48</td>
<td>.46</td>
<td>.35</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>.39</td>
<td>.33</td>
<td>.45</td>
<td>.40</td>
<td>.33</td>
</tr>
<tr>
<td>LIWC</td>
<td>AB</td>
<td>.38</td>
<td>.34</td>
<td>.45</td>
<td>.41</td>
<td>.42</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>.38</td>
<td>.36</td>
<td>.56</td>
<td>.47</td>
<td>.32</td>
</tr>
<tr>
<td>Combined</td>
<td>AB</td>
<td>.43</td>
<td>.33</td>
<td>.48</td>
<td>.40</td>
<td>.34</td>
</tr>
<tr>
<td></td>
<td>RF</td>
<td>.41</td>
<td>.36</td>
<td>.53</td>
<td>.47</td>
<td>.39</td>
</tr>
<tr>
<td>Baseline</td>
<td>–</td>
<td>.28</td>
<td>.26</td>
<td>.34</td>
<td>.32</td>
<td>.27</td>
</tr>
</tbody>
</table>
Success at Deception Detection Ability

• Baseline for ability to detect deception : 52%
  • AdaBoost, prosodic + LIWC features: 61%
  • Adding knowledge of gender, native language, personality: 65%
  • Useful features: ‘focuspresent’, speaking rate
• Currently unable to predict success at deception production above chance – but not yet using full set of lexical features
Other Questions

• Is laughter in deceptive speech produced differently from laughter in non-deceptive speech?
• Do people who entrain to their interlocutors detect deception better? Deceive better?
• Do acoustic-prosodic and lexical features that predict deception differ across cultures?
Next Class

• Emotional speech